IJLST

ISSN: 0974-5335
IJLST (2012), 5(1):1-7

Sequence Alignment of Triplex Capsid Protein of Human Herpes Simplex Virus

Vipan Kumar Sohpal[1], Apurba Dey[2], Amarpal Singh[3]
1 Department of Chemical & Bio Engineering BCET, Gurdaspur, Punjab INDIA
2 Department of Bio Technology NIT, Durgapur, West Bengal INDIA
3 Department of Electronics & Comm. Engg BCET, Gurdaspur, Punjab INDIA
Vipan752002@gmail.com, S_amarpal@yahoo.com

ABSTRACT:
Sequence comparison positioned at the centre of bioinformatics analysis. It is an important step toward structural and functional analysis of sequences. Pair wise sequence and multiple sequence alignment are the techniques of aligning the sequences on basis of database similarity. In first section this paper, we performed the pair wise alignment of viral capsid proteins of human herpes simplex virus (HHV) using heuristic and dynamic algorithm. In second part we implemented exact method, progressive alignment and iterative approach for multiple sequence alignment of viral protein data. The results from our experiments demonstrate that the multiple sequence alignment is more sensitive method than pair wise alignment can be used as an efficient computational platform for high performance sequence alignment applications. In later section of the paper we performed multiple sequence alignment of viral protein with hidden markov model (HMM).

Keywords: Pair wise & multiple sequence alignment, HHV capsid protein, heuristic and dynamic algorithm, progressive alignment and iterative approach, HMM

INTRODUCTION

Bioinformatics comprises of two sub-fields, first theme area is the development of computational tools and databases and second emerging area is the application of these tools in generating biological knowledge to better understand living systems. These computational tools are used in three areas namely molecular sequence analysis, molecular structural analysis, and molecular functional analysis. The areas of sequence analysis include sequence alignment, sequence data-base searching, motif and pattern discovery, gene and promoter finding, reconstruction of evolutionary relationships, and genome assembly and comparison. The most important and significant part of sequence analysis for comparison is sequence alignment. Sequence alignment is the technique by which sequences are compared by searching for common character patterns and establishing residue–residue correspondence among related sequences. Pair wise sequence alignment is the process of aligning two sequences and multiple sequence alignment is a natural extension of pair wise alignment, which is to align multiple related sequences to achieve optimal matching of the sequences. Although multiple sequence alignment has unique advantage because it reveals more biological information than pair wise alignments, but the higher computing time and memory reduce its merits. As a consequence, dynamic programming using Needleman–Wunsch cannot be applied for datasets of more than ten sequences. Progressive and iterative multiple sequence alignment approaches are most often used to overcome the limitation of dynamic programming.

Protein alignment is more informative than nucleotide alignment, because important relationships between related amino acids in an alignment can be accounted for using scoring systems. The human herpes simplex virus capsid proteins infect humans to cause a variety of illnesses including varicella, herpes zoster cancers, and even can cause brain inflammation. The viral genome of HHV cause infection through replication of genetic material facilitates to interact with glycoprotein and DNA maturation resultant infectious diseases. By analyzing the triplex capsid protein of human herpes simplex virus and other herpes simplex virus, it is possible to identify domains or motifs that are shared among a particular protein. These analyses of relatedness of protein are accomplished by alignment. Sequence alignment either progressive or heuristic requires high computational software and their availability. Dynamic programming using N-W and S-W algorithms, Blast, Clustal, Muscle have answers of the entire query for pair or multiple sequence alignment.

LITERATURE REVIEW

Jacek Blazewicz et al [1] implemented graphics processing unit using global and semi-global Needleman-Wunsch, and Smith-Waterman algorithms to construct the alignment from biological database. They presented the solution that performs the alignment of every given sequence pair, which is a required step for progressive multiple sequence alignment methods. Wang et al [2] compared the traditional affine gap penalty (AGP) and the bilinear gap penalty (BGP), with two profile-based variable gap penalty functions, on some well-established benchmark datasets. Robert C Edgar [3] compared the speed and accuracy of MUSCLE with CLUSTALW, Progressive POA and the MAFFT script. They observed MUSCLE-fast to be the fastest algorithm on all test sets, to achieve alignment accuracy and computational speed. Robert C. Edgar [4] put forward

MUSCLE, a new computer program for creating multiple alignments of protein sequences. Elements of the algorithm include fast distance estimation using kmer counting, progressive alignment using a new profile function. Landan & Graur [5] characterized pair wise and multiple sequence alignment (MSA) errors by comparing true alignments from simulations of sequence evolution with reconstructed alignments. Peris and Marzal [6] developed normalized global alignment score method to correct the length dependence of global alignments. Observation shows that normalized global alignment has a computational cost equivalent to 2.5 Needleman-Wunsch runs and a linear relationship with *Z*-score. Heringa [7] described three weighting schemes for improving the accuracy of progressive multiple sequence alignment methods: (1) global profile pre-processing (2) local pre-processing and (3) local–global alignment to improve alignment quality. Thompson et al [8] measured the sensitivity of the commonly used progressive multiple sequence alignment method and inculcated in Clustal W. Kumar S et al [9] developed the molecular evolutionary genetic software for sequence alignment and phylogenetic analysis. This software is user friendly for bioinformatics application. Mount [10] published reports that compared and analyzed the alignment effectiveness for different computational tool, along with varying algorithm. The focused area of his research was study the influence of alignment algorithm, amino acid scoring Matrix, and Gap Penalties on sequence alignment. Previous Work in this area has also included Clustal W is one of the best-known sequence alignment tools based on progressive approach. The main problem Clustal W is that Clustal W is that the initial pair wise alignments are fixed, and early errors cannot be corrected later, even if those alignments conflict with sequences added later. Myers and Miller [11] developed a linear-space version of Gotoh's algorithm, which accommodates affine gap penalties. Gotoh [12] covered important applications of multiple alignments for elucidation of the FESS relationships and expanding area of bioinformatics. Jimin [13] focused on methodologies and recent advances in the multiple protein sequence alignment field, with emphasis on the use of additional sequence and structural information to improve alignment quality. Needleman and Wunsch [14] performed protein sequence comparison using computer adaptable method through two-dimensional and pathways array. Team observed that maximum match is the largest number that would result from summing the cell values of every pathway. Tonges et al [15] developed a fast heuristic algorithm for multiple sequence alignment which provides near-to-optimal results for sufficiently homologous sequences. Zhang et al [16] worked upon blocking, chaining and flattening, that arise when computing a multiple-sequence alignment from given pair wise alignments. They developed practical algorithms which are effective for analyzing sequences containing internal repeats. Stephen et al [17] conclude that Blast a rapid local alignment search tool having simple and robust characteristics and an order of magnitude faster than existing sequence comparison tools of comparable sensitivity. This paper is devoted to the application of these bioinformatics tool to the sequence alignment [18] and [19]. The motivation for this work is implementation semi dynamic programming for pair and multiple sequence alignment and evaluation the performance of the resulting methods for score calculation on independent test sets of human herpes simplex virus. The effect of HMM on multiple sequence alignment also studied.

METHODS

There are various algorithms which are used for pair and multiple sequence alignment. We can access the relationship of any two protein sequence directly next to each other. One of the practical ways is BLAST tool. Another algorithm commonly used for global alignment is Needleman Wunsch approach.

Heuristic approach

Alignments using dynamic programming methods (S-W and N-W algorithm) are accurate and reliable, but too slow and impractical when computational resources are limited. Two popular local alignment algorithms have been developed that provide rapid alternatives to Smith–Waterman namely FASTA (Pearson and Lipman, 1988) and BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990). The main merits of these algorithms, requires less time to perform an alignment. The time saving occurs because these methods restrict the search by scanning a database for likely matches before performing more rigorous alignments. These are heuristic algorithms that sacrifice some sensitivity in exchange for speed and are not guaranteed to find optimal alignments.

Dynamic programming

One of the first and most important algorithms for aligning two protein sequences was described by Saul Needleman and Christian Wunsch (1970), with subsequent modifications by Sellers (1974), Gotoh (1982), and others. This algorithm is important because it produces an optimal alignment of two protein or DNA sequences, even allowing the introduction of gaps. The result is optimal, but nonetheless not all possible alignments need to be evaluated. The Needleman–Wunsch algorithm is an example of dynamic programming in which the optimal alignment is identified by reducing the problem to a series of smaller alignments on a residue-by-residue basis.

Exact approach

Exact methods for multiple sequence alignment employ dynamic programming, although the matrix is

multidimensional rather than two-dimensional. The goal is to maximize the summed alignment score of each pair of sequences. Exact methods generate optimal alignments but are not feasible in time or space for more than a few sequences. For N sequences, the computational time that is required is O ($2^N L^N$) where N is the number of sequences and L is the average sequence length. An exact multiple sequence alignment of more than four or five average sized proteins would consume prohibitively too much time.

Progressive approach

This method was popularized by Feng and Doolittle (1987, 1990). Progressive approach involve a strategy entails calculating pair wise sequence alignment scores between all the proteins being aligned, then beginning the alignment with the two closest sequences and progressively adding more sequences to the alignment. A benefit of this approach is that it permits the rapid alignment of even hundreds of sequences. A major limitation is that the final alignment depends on the order in which sequences are joined. Thus, it is not guaranteed to provide the most accurate alignments, which reduce its authenticity.

Iterative approach

Iterative approaches can overcome error occurs in the alignment process that reduce the limitation of progressive approach. Iterative refinement can search for more optimal solutions stochastically or by systematically extracting and realigning sequences from an initial profile that is generated. MUSCLE (Robert Edgar 2004a, 2004b) has become popular iterative approach because of its accuracy and its exceptional speed, especially for multiple sequence alignments involving large numbers of sequences and operates in a successive three stages.

Proposed Methodology

A proposed change in the alignment technique is sum of heuristic, dynamic and progressive approach. It involves various stages for accuracy and less time computing process. In pair wise alignment of human herpes simplex viral protein, we performed Blastp operation to evaluate the alignment parameters. Those parameters used to evaluate the optimal scoring matrices using dynamic programming. Similarly multiple sequence alignment performed using progressive and iterative approach, their resultant parameters used for dynamic programming.

BENCHMARK DATA

Reference data are available with number of protein database. The proteomic data for sequence alignment of triplex capsid protein of herpes virus were obtained from the Uniprot KB. UniProt database has larger coverage than any one of the three databases (SWISS-PROT, TrEMBL, and PIR) while at the same time maintaining the original SWISS-PROT feature of low redundancy, cross-references, and a high quality of annotation. Data accessed from UniProt resources having accession P32888, P22486 and P89461 of HHV-1 and HHV-2 strains. There are 465 pair-wise reference alignments by Muscle using MEGA. Data sets are marked as model 1, 2 and 3 of different strains according to accession number having sequence length 466.

RESULTS

Blastp search operation run by pasting sequence [accession number of human herpes simplex virus as mentioned in benchmark data] into BLAST input box against the non redundant (nr) database at substitution matrix of Blosum62. A typical Blastp output reports both E values and scores represented in table1. Alignment score is important parameter to detect the similarity in form of statistics. The Blastp scores are represented into raw and bit scores. Raw scores are depending upon substitution matrix, while bit score is calculated from the raw score by normalizing with the statistical variables that define a given scoring system.

$$E = K * m * n * e^{-\lambda * S}$$

E refers to the expect value, which is the number of different alignments with scores equivalent to or better than S that are expected to occur by chance in a database search. E value score is function of λ (scales the scoring system), length of the query sequence & the length of the database (K is a scaling factor for the search space). These parameters (K and λ) for viral capsid protein of ungapped and gapped shown in table 2

Dynamic programming with Needleman–Wunsch algorithm is method in which the optimal alignment is identified by reducing the problem to a series of smaller alignments on a residue-by-residue basis. We use the Needleman–Wunsch approach to global sequence alignment in three phases. Phase 1 involve setting up a matrix with alignment parameters obtained in section 1.Second phase meant scoring the matrix, and third phase identifying the optimal alignment on the basis of score. We used the BLOSUM matrices less divergent database of human herpes simplex virus. Figure1 shows the results of optimizing model 3 on fixed alignment parameter gap costs (10, 2) obtained from heuristic method (Blastp). The results are qualitatively similar on three models. The results show model3 is the best for all the specified matrices range of closely related species. Moreover the differences in bit scores in three species are exceptionally small and having less than 0.5% variation throughout the substitution matrices.

ISSN: 0974-5335
IJLST (2012), 5(1):1-7

Exact and progressive method for multiple sequence alignment created in a series of steps using Clustal W. The algorithm first selects the two most closely related sequences from the guide tree and creates a pair wise alignment. P32888 and P23210 are aligned. The next sequence is either added to the pair wise alignment to generate a profile. Process of addition of new sequence continues until the root of the tree is reached, and all sequences have been aligned as shown in table 3. The highest score observed in alignment of sequence P23210 and P22486.While in the alignment of three closely related capsid proteins, we observed that a highly conserved P32888 is aligned (Fig. 2) as is a P23210 and P22486 that coordinates. The result in figure 2 indicates that capsid protein of closely related human herpes simplex virus, the conservation is so high that there are no gaps in alignment of proteins.

Iterative methods compute a suboptimal solution using a progressive alignment strategy, and then modify the alignment using dynamic programming or other methods until a solution converges. Muscle also works in different

Successive stages for alignment process. Muscle detects the similarities using a triangular distance matrix, and then constructs a rooted tree using UPGMA or neighbor-joining. In this work we accessed the Muscle through Molecular evolutionary genetic analysis tool. The alignment of triplex capsid protein of human herpes simplex virus observed at specific gap penalties (10, 2) using Neighbor joining clustering method. The alignments of three closely related viral proteins using Clustal W and Muscle show a different result from each other. In the Fig 3 there are only few variation in alignment but reflects a more compact overall alignment. This program still shows to the highly conserved regions

Table 2 Output of Blastp of Karlin–Altschul statistics are provided and can be used to relate scores to expect values.

Parameters	Ungapped	Gapped
λ	0.321418	0.267
K	0.137192	0.041
H	0.43678	0.14

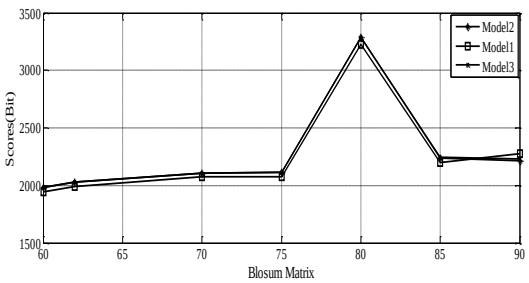

Figure 1 Global pairwise alignment using a series Blosum matrix. Closely related viral capsid protein were aligned using a series of Blosum (x-axis) and bit scores were measured (y-axis)

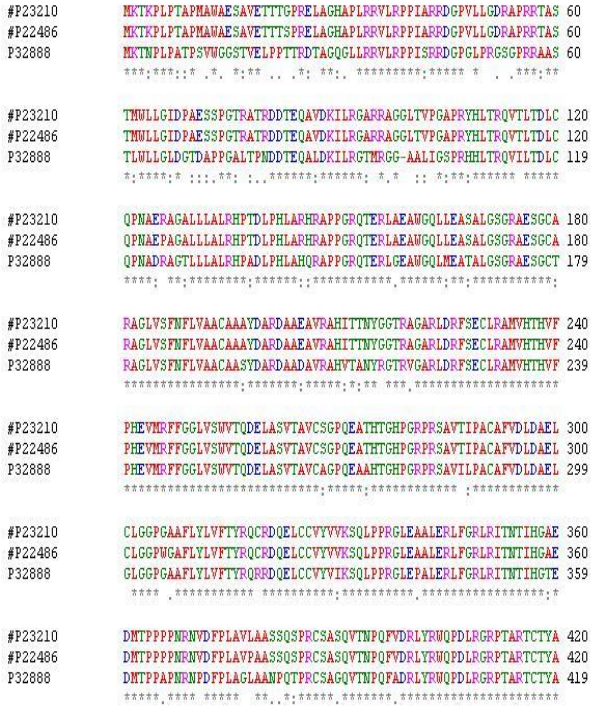

Figure 2 Multiple sequence alignment of three viral capsid proteins. The output is from Clustal W 2 using the progressive alignment approach

Table: 3 A series of pair-wise alignment is generated for three viral capsid protein of human herpes simplex virus using clustalw2 (2.1)

SeqA	Name	Length	SeqB	Name	Length	Score
1	P32888	465	2	#P23210	466	81.0
1	P32888	465	3	#P22486	466	80.0
2	#P23210	466	3	#P22486	466	98.0

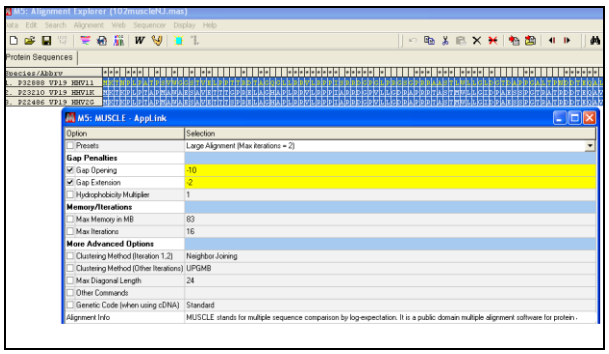

Figure 3 Multiple sequence alignment using in-built Muscle tool in MEGA at specific gap penalties (10, 2) and Neighbor joining clustering method

Table 1 A Blastp output includes a list of database sequences that match the query. The score and E value for each alignment are also provided. The best matches at the top of the list have large score and corresponding E values.

Accession	Description	Max score	Total score	Query coverage	E value
P32888.1	RecName: Full=Triplex capsid protein VP19C; AltName: Full=Virion protein UL38	936	936	100%	0.0
P17586.1	RecName: Full=Triplex capsid protein VP19C; AltName: Full=Virion protein UL38	926	926	100%	0.0
P22486.1	RecName: Full=Triplex capsid protein VP19C	736	736	100%	0.0
P89461.1	RecName: Full=Triplex capsid protein VP19C; AltName: Full=Capsid protein VP19C	722	722	100%	0.0
P28935.1	RecName: Full=Triplex capsid protein 22; AltName: Full=Capsid protein VP19C	328	328	77%	4e-107
Q6S6P9.1	RecName: Full=Triplex capsid protein 22; AltName: Full=Capsid protein VP19C	326	326	77%	3e-106
P09276.1	RecName: Full=Triplex capsid protein VP19C	259	259	76%	3e-80
Q9E6N1.1	RecName: Full=Triplex capsid protein VP19C; AltName: Full=Virion protein UL38	252	252	78%	2e-77
Q6UDJ3.1	RecName: Full=Triplex capsid protein VP19C; AltName: Full=Virion protein UL38	112	112	78%	1e-25
A6VIP7.1	RecName: Full=TATA-box-binding protein; AltName: Full=Box A-binding protein; Short=BAP; AltName: Full=TATA sequence-binding protein; Short=TBP; AltName: Full=TATA-box factor	37.0	37.0	14%	0.13
A6URP5.1	RecName: Full=TATA-box-binding protein; AltName: Full=Box A-binding protein; Short=BAP; AltName: Full=TATA sequence-binding protein; Short=TBP; AltName: Full=TATA-box factor	36.2	36.2	15%	0.28
Q9P9I9.1	RecName: Full=TATA-box-binding protein; AltName: Full=Box A-binding protein; Short=BAP; AltName: Full=TATA sequence-binding protein; Short=TBP; AltName: Full=TATA-box factor; AltName: Full=aTBP	35.0	35.0	14%	0.60
Q8G7Y3.1	RecName: Full=Undecaprenyl pyrophosphate synthase; Short=UPP synthase; AltName: Full=Di-trans,poly-cis-decaprenylcistransferase; AltName: Full=Undecaprenyl diphosphate synthase; Short=UDS	35.4	35.4	15%	0.72
A9A840.1	RecName: Full=TATA-box-binding protein; AltName: Full=Box A-binding protein; Short=BAP; AltName: Full=TATA sequence-binding protein; Short=TBP; AltName: Full=TATA-box factor	34.3	34.3	14%	1.1
Q8WU79.1	RecName: Full=Stromal membrane-associated protein 2; AltName: Full=Stromal membrane-associated protein 1-like	32.7	32.7	12%	5.7
Q7TN29.1	RecName: Full=Stromal membrane-associated protein 2; AltName: Full=Stromal membrane-associated protein 1-like	32.7	32.7	12%	5.8
Q15751.2	RecName: Full=Probable E3 ubiquitin-protein ligase HERC1; AltName: Full=HECT domain and RCC1-like domain-containing protein 1; AltName: Full=p532; AltName: Full=p619	33.1	33.1	12%	6.0

Table 4 A pair wise alignment viral capsid protein is generated using HMMER

Family		Description	Start	End	Alignment		Model		Bias	Accuracy	Bit Score	Domain E-values	
Id	Accession						Start	End				Ind.	Cond.
Herpes_VP 19C	PF03327.8	Herpes virus capsid shell Protein VP 19C	148	464	149	464	2	270	0.0	0.99	297.0	8.1e-89	6.6e-93

Pfam is one of the most comprehensive databases of protein families. It is a compilation of both multiple sequence alignments and profile HMMs of protein families. Pfam-A is collection of protein families in the form of multiple sequence alignments and profile HMMs. We used HMMER software is used to perform searches. Sequences in Pfam-A are grouped in families, assigned stable accession numbers as shown in table 4. The observed results indicate that accuracy on higher side while an E value on lower side indicates highest similarity.

DISCUSSION

Blastp display maximum score, total score, query coverage and E value in decreasing order, highest from the top and lowest at terminating position. Maximum score observed for human herpes simplex virus for given accession number (P32888.1) with nr database is 936 having query coverage 100%. The range of maximal score varied from 936 to 33.1. The variation in maximal score and total score function query converge. It (query coverage) depends upon length coverage of the input query sequence by different HSPs from the same database sequence. The low query

converges responsible for low score in Blastp output. The minimum query coverage 12% for accession number Q15751.2.This indicates that protein database of query and nr database have quite difference due to variation in capsid protein of virus. On other hand E Value (Expect Value) describes the likelihood that a sequence with a similar score will occur in the database by chance. Alignments having a lowest E value (P28935.1, Q6S6P9.1, P09276.1 and Q9E6N1.1) of meaning that a sequence with a similar score is very unlikely to occur simply by chance. Expected score mathematically depend upon (λ, K, H), gapped and ungapped alignment method. Karlin–Altschul statistics values (λ) of ungapped and gapped alignments are respectively .3214, 0.267. The K value variation is more than 68% in ungapped and gapped alignment of human herpes simplex virus. Similarly the variation in H value more than 60% for ungapped and gapped method. The heuristic method performs the optimum at gap costs (10, 2).

Fig. 1 shows the results of optimizing sequence (extension penalty e (2.0) with fixed gap-open penalty g = 10) on the Emboss software using inbuilt Blosum substitution matrix .The results indicated in Fig 2. is qualitatively and quantitative similar on the three sets, giving confidence that they indicate general trends rather than artifacts of benchmark construction, of overtraining or of significantly suboptimal local maxima. This is further confirmed by alignment score (Fig.1), which again gives similar, results, as would be expected in substitution matrices. The alignment score exists in range of 2010-3350 in all three models. In all the tests (optimal score) reported at Blosum 80. The highest % of alignment similarity observed in all three sequences of capsid proteins in dynamic alignment technique.

Multiple sequence alignment of human herpes simplex virus shown in Fig 2, which reflect global alignment for all the three sequences. The sequence having accession number of P32888, #P23210 and #P22486 aligned and score observed in range of 80-98. The sequences (2:3) have highest score, (1:3) shows lowest score, while (1:2) score is intermediate. It indicates sequence (2:3) have common ancestor or less divergence. On other hand sequence (1:3) shows high divergence among three sequences. The variation in upper and lower score is 17%. Muscle multiple sequence alignment can be used through MEGA. The result observed under constraint of gap opening and extension under neighbor joining method in Fig 3. The output of MEGA for muscle platform shows highly conserved sequence. The start/end of the MEA alignment of this Herpes VP 19C with respect to the profile HMM, which directly relates to the query sequence shown in table 4. The target envelope lies between 149-464.The low bias composition corrections for true positive homologous sequences. The accuracy ranges 0.99 indicating a completely reliable alignment according to the model. The independent and conditional E value is significant to observe homology

CONCLUSIONS

On line sequence alignment is better option, but these methods and algorithms have some limitation. This paper focused on combination of bioinformatics tool with modified algorithms. Overall Blastp is basic tool and result of this used for base of pair wise dynamic programming. The output of dynamic method used for multiple sequence alignment using progressive method for motif and domain. The MSA with HMM used to detect the homology on basis of accuracy range

REFERENCES

[1] Blazewicz Jacek, Frohmberg Wojciech, Kierzynka Michal, Pesch Erwin, Wojciechowski Pawel. (2011) Protein alignment algorithms with an efficient backtracking routine on multiple GPUs, BMC Bioinformatics 12 (1): 181-192

[2] Chuan Wang, Ren-Xiang Yan, Xiao-Feng Wang, Jing-Na Si, Ziding Zhang. (2011) Comparison of linear gap penalties and profile-based variable gap penalties in profile–profile alignments Computational Biology and Chemistry, 35(5):308-318

[3] Edgar, R.C (2004) MUSCLE: a multiple sequence Alignment method alignment method with reduced time and space complexity BMC Bioinformatics, 11(5) doi:10.1186/1471-2105-5-113

[4] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5) :1792-1797

[5] Giddy Landan, Dan Graur (2009) Characterization of pair wise and multiple sequence alignment errors Gene, 441(1-2): 141-147

[6] Guillermo Peris, Andres Marzal (2011) Normalized global alignment for protein sequences, Journal of Theoretical Biology, 291:22-28

[7] Jaap Heringa (2002) Local weighting schemes for protein multiple sequence alignment, Computers & Chemistry, 26(5):459-477

[8] JD Thompson, DG Higgins (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix, Nucleic Acids Research, 22:4673-4680

[9] Kumar S, Nei M, Dudley J, Tamura K. (2008) MEGA: biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinformatics 9:299-306

[10] Mount D.W (2008). Studies of varying alignment algorithm, amino acid scoring matrix and gap penalties. CSH Protocols doi: 10.1101/pdb.ip60

[11] Myers EW, Miller W (1988) Optimal alignments in linear space, Computer Application Bioscience 4:11-17.

[12] Osamu Gotoh (1999) Multiple sequence alignment algorithms and application, Advance in Biophysics, 36: 159-206

ISSN: 0974-5335
IJLST (2012), 5(1):1-7

[13] Pei J (2008) Multiple protein sequence alignment Current Opinion in Structural Biology 18 (3):382-386

[14] Saul B. Needleman, Christian D. Wunsch (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology 48 (3): 443-453

[15] Udo Tonges, Soren W. Perrey, Jens Stoye, Andreas W.M. Dress (1996) A general method for fast multiple sequence alignment Gene 172(1):33-41

[16] Zhang Z., B. He, W. Miller (1996) "Local multiple alignment via sub graph" Discrete Applied Mathematics, Volume 71, Issues 1-3, Pages 337-365

[17] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, David J. Lipman (1990) Basic local alignment search tool, Journal of Molecular Biology, 215(3): 403-410

[18] Sohpal, V.K., Dey, A. and Singh, A (2010) MEGA biocentric software for sequence & phylogenetic analysis: a review, Int. J. Bioinformatics Research and Applications, 6(3): 230–240

[19] Sohpal, V.K., Dey, A. and Singh, A (2010) Sequence alignment and Phylogenetic Analysis of Human Herpes Simplex Virus (HHV) using Bioinformatics Tool: A Review, Int. J. Computational Biology & Drug Design 3(1) 68-88